第27卷 第2期　工 程 数 学 学 报　Vol. 27 No. 2
2010年04月　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　Apr. 2010

文章编号:1005-3085(2010)02-0342-05

不完备信息系统确定性和集对联系度的粗集拓展模型*

李长清[1]， 李克典[2]， 李进金[2]

(1- 汕头大学理学院，汕头 515063； 2- 漳州师范学院数学与信息科学系，漳州 363000)

摘　要: 本文把集对分析的思想融入不完备信息系统确定性理论之中，得到了更广泛的拓展模型，它是通过调整参数，达到对系统的理想分类。这种模型既保留了已有的拓展模型的优点，又克服了它们的局限性，为处理不完备信息系统提供了一种有效的方法。

关键词: 不完备信息系统；联合确定率；集对分析；联系度

分类号: **AMS**(2000) 68T10; 68T37　**中图分类号**: TP18; TP311　**文献标识码**: A

1　引言

粗集理论[1] 是近年来发展起来的一种新的处理不确定性和不完整知识的数学工具，已被广泛应用于决策与分析、模式识别和数据挖掘[2,3] 等领域。经典的粗集理论处理的是完备的信息系统，而在现实应用中，由于对数据获取的限制，使得大量的数据缺失或不确定，从而信息系统都不完备。因此，对经典的粗集理论进行必要的拓展成为人们研究的一个热门课题，现已有了许多广泛拓展。主要有基于容差关系[4]、相似关系[5] 和限制容差关系[6] 的粗集拓展模型，它们在不完备信息系统的处理中各自具有明显的特征，但这些拓展模型也都存在一定的局限性。

赵克勤教授提出的集对分析方法 (SPA)[7] 是一种用于研究集合之间相互联系的新理论，现已得到广泛的应用。黄兵[8] 针对以上几种粗集拓展模型的局限性，运用集对分析方法对它们进行改进，得到更好的拓展模型。随后，刘富春[9] 做进一步的拓展。但以上进一步拓展的模型仍存在局限性，它们只能解决某一层次的问题，对于复杂的不完备信息系统不能很好的处理。本文把集对分析的思想融入不完备信息系统确定性理论中，通过引入参数对以上的粗集模型进一步拓展，得到了更广泛的拓展模型。它是利用调整参数，达到对系统的理想分类，既保留了前面几种拓展模型的优点，又克服了它们的局限性，最后举一个实例进行分析。

2　几种已有粗集拓展模型

给定不完备信息系统 $S=\langle U, A\rangle$，其中 U 是论域，A 是有限非空属性集，对任意的 $a\in A$，有 $a:U\to V_a$，V_a 为 a 的值域，$*$ 表示不确定或已丢失的属性值。

定义 2.1[4]　设 $\langle U, A\rangle$ 是不完备信息系统，$B\subseteq A$，容差关系 T 定义为

$$T_B=\big\{(x,y)\in U\times U \mid \forall\, a\in B,\ a(x)=*\vee a(y)=*\vee a(x)=a(y)\big\}.$$

令 $T_B(x)=\{y\in U \mid (x,y)\in T_B\}$。

收稿日期: 2008-03-10.　**作者简介**: 李长清 (1979年3月生)，男，博士. 研究方向：粗集与拓扑学理论及其应用.

***基金项目**: 国家自然科学基金 (10671173; 10971195; 10971186)；福建省自然科学基金 (2008F5066).

定义 2.2[5]　设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$，非对称相似关系 S 定义为

$$S_B = \big\{(x,y) \in U \times U \mid \forall\, a \in B,\ a(x) = * \vee a(x) = a(y)\big\}.$$

定义 2.3[6]　设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$，限制容差关系 L 定义为

$$L_B = \{(x,y) \in U \times U \mid \forall\, a \in B(a(x) = a(y) = *)$$
$$\vee(P_B(x) \cap P_B(y) \neq \emptyset \wedge \forall\, a \in B((a(x) \neq * \wedge a(y) \neq *) \to a(x) = a(y)))\},$$

其中 $P_B(x) = \{a \in B \mid a(x) \neq *\}$。令 $L_B(x) = \{y \in U \mid (x,y) \in L_B\}$。

容差关系和非对称相似关系是对完备信息系统下等价关系的两种拓展，它们是两个极端情形，前者相对宽松，而后者又相对紧凑。限制容差关系刚好介于这两者之间，比较好地体现了优越性。但它仍存在不足，对于个体之间只要有一个属性值不同 (如 $x = (0,1,2,\cdots,99)$ 与 $y = (1,1,2,\cdots,99)$) 就看作不同，被划在不同类中，这会导致知识库划分过细，不利于对系统的处理。文献 [9] 定义的集对 (α,β) 容差关系，通过对集对联系度的调节，达到对系统灵活有效地处理，克服了以上不足。

定义 2.4[9]　设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$。对任意的 $x, y \in U$，x 与 y 在 B 下的集对联系度 $U_B(x,y)$ 定义为

$$U_B(x,y) = S_1 + S_2 i + S_3 j,$$

这里 $S_1 = |M(x,y)|/n$ 称为 x,y 在 B 下的同一度，$S_2 = |N(x,y)|/n$ 称为 x,y 在 B 下的对立度，$S_3 = |K(x,y)|/n$ 称为 x,y 在 B 下的差异度，$n = |B|$。其中

$$M(x,y) = \big\{a \in B \mid a(x) \neq * \wedge a(y) \neq * \wedge a(x) = a(y)\big\},$$

$$N(x,y) = \big\{a \in B \mid a(x) \neq * \wedge a(y) \neq * \wedge a(x) \neq a(y)\big\},$$

$$K(x,y) = \big\{a \in B \mid a(x) = * \vee a(y) = *\big\}.$$

定义 2.5[7]　设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$。集对 (α,β) 容差关系 $SP(\alpha,\beta)$ 定义为

$$SP_B(\alpha,\beta) = \big\{(x,y) \in U \times U \mid U_B(x,y) = S_1 + S_2 i + S_3 j \wedge S_1 > \alpha \wedge S_2 \leq \beta\big\} \cup I,$$

其中 $I = \{(x,x) \mid x \in U\}$。令 $SP_B(\alpha,\beta)(x) = \{y \in U \mid (x,y) \in SP_B(\alpha,\beta)\}$。

容差关系，限制容差关系和集对 (α,β) 容差关系对属性值有明确相同的个体都有各自的归属，而对没有一个明确相同的个体归属都存在片面性。例如对 $x = (1,2,*,\cdots,*)$，$y = (*,*,*,\cdots,*)$，$z = (*,*,*,\cdots,*)$，容差关系认为 x, y, z 不可分辨，限制容差关系认为 x 与 y 可分辨，y 与 z 不可分辨，集对容差关系认为 x, y, z 两两可分辨，这就与实际不相符。鉴于以上局限性，以下结合集对分析思想与对象属性值联合确定率这个概念，提出一种不完备信息系统确定性和集对联系度的粗集拓展模型。

3　确定性和集对联系度的粗集拓展模型

定义 3.1[10]　设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$，对任意的 $x, y \in U$，对象 x,y 属性值联合确定率 $F(x,y)$ 定义为

$$F(x,y) = \begin{cases} 1, & x = y, \\ \dfrac{|P_B(x)| + |P_B(y)|}{2|B|}, & x \neq y. \end{cases}$$

其中 $P_B(x) = \{b \in B \mid b(x) \neq *\}$，易知 $0 \leq F(x,y) \leq 1$。

定义3.2 设 $\langle U, A\rangle$ 是不完备信息系统，$B \subseteq A$，$0 \leq \alpha,\beta,\gamma \leq 1$，确定集对容差关系 $J(\alpha, \beta, \gamma)$ 定义为

$$J_B(\alpha,\beta,\gamma) = \big\{(x,y) \in U \times U \mid U_B(x,y) = S_1 + S_2 i + S_3 j$$

$$\wedge((S_1 > \alpha \wedge S_2 \leq \beta) \vee (S_3 = 1 \wedge F(x,y) \geq \gamma))\big\} \cup I,$$

这里 $I = \{(x,x) \mid x \in U\}$。显然 $J(\alpha,\beta,\gamma)$ 满足自反性和对称性，但不满足传递性。x 的确定集对容差类为 $J_B(\alpha,\beta,\gamma)(x) = \{y \in U \mid (x,y) \in J_B(\alpha,\beta,\gamma)\}$。

对任意的 $X \subseteq U$，X 的下近似 (X_B^J) 和上近似 (X_J^B) 为

$$X_B^J = \big\{x \in U \mid J_B(\alpha,\beta,\gamma)(x) \subseteq X\big\}, \quad X_J^B = \big\{x \in U \mid J_B(\alpha,\beta,\gamma)(x) \cap X \neq \emptyset\big\}.$$

定理3.1 设 $S = \langle U, A\rangle$ 为不完备信息系统，$J_A(\alpha,\beta,\gamma)$ 为确定集对容差关系，则

1) 当 $\alpha = \beta = \gamma = 0$ 时，$J_A(\alpha,\beta,\gamma) = T_A$;

2) 当 $0.5 < \gamma \leq 1$ 时，$J_A(\alpha,\beta,\gamma) = SP_A(\alpha,\beta)$;

3) 若对任意的 $x \in U$，$P(x) \neq \emptyset$，则当 $\alpha = \beta = 0$，$0.5 < \gamma \leq 1$ 时，$J_A(\alpha,\beta,\gamma) = L_A$。

此定理表明，确定集对容差关系是容差关系、限制容差关系以及集对 (α,β) 容差关系的推广。

设 $S = \langle U, A\rangle$ 是不完备信息系统，$X, Y \subseteq U$，$B \subseteq A$。从定义3.2可以直接得到确定性和集对联系度的粗集上、下近似的如下性质。

性质3.1 $X_B^J \subseteq X \subseteq X_J^B$，$\emptyset_J^B = \emptyset$，$U_B^J = U$。

性质3.2 若 $X \subseteq Y$，则 $X_B^J \subseteq Y_B^J$，$X_J^B \subseteq Y_J^B$。

性质3.3 $(X \cap Y)_B^J = X_B^J \cap Y_B^J$，$(X \cup Y)_J^B = X_J^B \cup Y_J^B$。

性质3.4 $(X \cap Y)_J^B \subseteq X_J^B \cap Y_J^B$，$(X \cup Y)_B^J \supseteq X_B^J \cup Y_B^J$。

性质3.5 $(X_B^J)_B^J \subseteq X_B^J \subseteq (X_B^J)_J^B \subseteq X \subseteq (X_J^B)_B^J \subseteq X_J^B \subseteq (X_J^B)_J^B$。

性质3.6 设 $X, Y \subseteq U$，则 $X \subseteq Y_B^J \Leftrightarrow X_J^B \subseteq Y$。

4 实例分析

设不完备信息系统 $S = \langle U, A\rangle$，其中论域 $U = \{x_1,x_2,x_3,x_4,x_5,x_6\}$ 表示6个家庭调查对象，属性集 $A = \{a_1,a_2,a_3,a_4,a_5,a_6,a_7,a_8\}$，$a_1$ 表示住房大与小，a_2 表示房价贵与便宜，a_3 表示住房装修与未装修，a_4 表示住房周边环境好与坏，a_5 表示住房楼层高与低，a_6 表示住房质量好与坏，a_7 表示住房新与旧，a_8 表示住房远与近，$V_i = \{1,0\}\,(i = 1,2,\cdots,8)$，如表1。

下面用容差关系分析这个信息系统。

$$T_A(x_1) = T_A(x_2) = \{x_1,x_2,x_3,x_4,x_6,\},$$

$$T_A(x_3) = T_A(x_4) = T_A(x_6) = \{x_1,x_2,x_3,x_4,x_5,x_6\},$$

$$T_A(x_5) = \{x_3,x_4,x_5,x_6\}.$$

在这种分类中，x_2 只有两个属性值不是空值，x_3 属性值都是空值，认为它们不可分辨导致了系统划分过粗，不利于对系统的处理，还有如 x_2 与 x_6，x_3 与 x_6 等都有类似情况。

表 1:　住房条件数据表

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
x_1	1	1	0	1	1	1	*	0
x_2	*	*	0	*	*	*	*	0
x_3	*	*	*	*	*	*	*	*
x_4	1	*	*	*	*	1	0	0
x_5	1	0	1	0	0	1	0	0
x_6	*	*	*	*	*	*	*	*

下面用限制容差关系分析这个信息系统。

$$L_A(x_1) = L_A(x_2) = \{x_1, x_2, x_4\}, \quad L_A(x_3) = L_A(x_6) = \{x_3, x_6\},$$

$$L_A(x_4) = \{x_1, x_2, x_4, x_5\}, \quad L_A(x_5) = \{x_4, x_5\}.$$

在这种分类中，x_2 只有两个属性值不是空值，x_3 属性值都是空值，认为它们可分辨，而 x_3 与 x_6 都是空值却认为不可分辨，这与实际很不相符。

最后，用集对 (α, β) 容差关系分析这个信息系统，取 $\alpha = \frac{1}{4}$，$\beta = \frac{1}{8}$。

$$SP_A(x_1) = \{x_1, x_4\}, \quad SP_A(x_2) = \{x_2\}, \quad SP_A(x_3) = \{x_3\},$$

$$SP_A(x_4) = \{x_1, x_4, x_5\}, \quad SP_A(x_5) = \{x_4, x_5\}, \quad SP_A(x_6) = \{x_6\}.$$

在这种分类中，x_5 属性值都不是空值，x_3 属性值都是空值，认为它们可分辨导致了系统划分过细，也不利于对系统的处理。

与以上对比，用确定集对容差关系分析这个信息系统，取 $\alpha = \frac{1}{4}$，$\beta = \frac{1}{8}$，$\gamma = \frac{7}{16}$。

$$J_A(x_1) = \{x_1, x_3, x_4, x_6\}, \quad J_A(x_2) = \{x_2\}, \quad J_A(x_3) = \{x_1, x_3, x_5\},$$

$$J_A(x_4) = \{x_1, x_4, x_5\}, \quad J_A(x_5) = \{x_3, x_4, x_5, x_6\}, \quad J_A(x_6) = \{x_1, x_5, x_6\}.$$

用这种关系分类，更为实际有效且完全克服了以上几种关系的局限性。

5　结束语

本文把集对分析的思想融入不完备信息系统确定性理论中，得到更广泛的不完备信息系统的粗集拓展模型。这种模型可根据实际应用的需要，对参数 α, β, γ 合理的调节，从而达到对更广泛的不完备信息系统的有效处理，既保留了已有几种拓展模型的优点，又克服了它们的局限性。

参考文献：

[1] Pawlak Z. Rough sets[J]. International Journal of Computer Information Science, 1982, 11: 341-356

[2] Pawlak Z, Busse J G, slowinski R, *et al.* Rough sets[J]. Communications of the ACM, 1995, 38(11): 89-95

[3] 张文修，吴伟志，梁吉业等. 粗集理论与方法[M]. 北京：科学出版社，2001
Zhang W X, Wu W Z, Liang J Y, *et al.* Theory and Method of Rough Sets[M]. Beijing: Science Press, 2001
[4] Kryszkiewicz M. Rough set approach to incomplete information system[J]. Information Sciences, 1998, 11(2): 39-49
[5] Stefanowski J, Tsoukias A. On the extension of rough sets under incomplete information[C]// Proc of the 7th Int'l Workshop on New Directions in Rough Set, Data Mining, and Granular soft computing, Berlin: Springer-Verlag, 1999: 73-81
[6] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展，2002, 39(10): 1238-1243
Wang G Y. Extension of rough set under incomplete information systems[J]. Journal of Computer Research and Development, 2002, 39(10): 1238-1243
[7] 赵克勤. 集对分析及其初步应用[M]. 杭州：浙江科学出版社，2000
Zhao K Q. Set Pair Analysis and its Primary Application[M]. Hangzhou: Zhejiang Science Press, 2000
[8] 黄兵，周献中. 基于集对分析的不完备信息系统粗糙集模型[J]. 计算机科学，2002, 29(7): 1-3
Huang B, Zhou X Z. Rough set model based on set pair analysis in incomplete information system[J]. Computer Science, 2002, 29(7): 1-3
[9] 刘富春. 变集对联系度的扩充粗糙集模型及其属性约简[J]. 计算机科学，2006, 33(3): 185-187
Liu F C. Variable set-pair connectivity extension of rough set model and its attributes reduction[J]. Computer Science, 2006, 33(3): 185-187
[10] 林仁炳，黄英. 不完备信息系统中变精度粗糙集模型[J]. 浙江传媒学院学报，2006, 1: 70-71
Lin R B, Huang Y. Variable precision rough set model in incomplete information system[J]. Journal of Zhejiang Institute of Media, 2006, 1: 70-71

Extension of Rough Sets Based on Determinism of Incomplete Information Systems and Set-pair Connection Degree

LI Chang-qing[1], LI Ke-dian[2], LI Jin-jin[2]

(1- College of Sciences, Shantou University, Shantou 515063; 2- Department of Mathematics and Information Science, Zhangzhou Normal College, Zhangzhou 363000)

Abstract: In view of the limitations in the existing extensions of rough set, this paper proposes a novel extension. By unifying determinism of incomplete information systems and set-pair analysis, we attain a better classification. The extension provides an effective method for incomplete information systems.

Keywords: incomplete information systems; union determine rate; set-pair analysis; connection degree

Received: 10 Mar 2008. **Accepted:** 16 July 2008.
Foundation item: The National Natural Science Foundation of China (10671173; 10971195; 10971186); the Natural Science Foundation of Fujian Province (2008F5066).